A distinct low lung function trajectory from childhood to the fourth decade of life

Cristine E. Berry MD, Dean Billheimer PhD, Isaac C. Jenkins MS, Zhenqiang J. Lu PhD, Debra A. Stern MS, Lynn B. Gerald PhD, Tara F. Carr MD, Stefano Guerra MD, Wayne J. Morgan MD, Anne L. Wright PhD, and Fernando D. Martinez MD

ONLINE DATA SUPPLEMENT

ONLINE SUPPLEMENT

Methods Supplement

We modeled the pre-bronchodilator FEV_1 to FVC ratio using a heterogeneous linear mixed effects model, adjusting for sex, in order to identify latent classes of lung function trajectory.[1,2] This model assumes that the random slope and intercept components for each participant are distributed as a mixture of Gaussians. When applied to longitudinal data, this allows us to detect latent classes of lung function trajectories as well as individual probabilities of class membership. This is done by pre-specifying the number of latent classes, which determines the number of Gaussian components in the mixture model. The mixture parameter is then interpreted as the probability of belonging to one of the latent classes. The results can be interpreted discretely (i.e. the highest probability determines the class membership) or on a continuous spectrum (i.e. we have a range for how certain we are a participant belongs to each class). We fit the model with one to three latent classes and used AIC for model selection. Note that latent class characteristics and class membership of participants are data-driven and not the result of a supervised classification.

Version 3.1.2 of the statistical programming language R was used for data exploration and analyses [@r]. The R package lcmm was used to fit the models. We used the package's hlme function, which utilizes a Marquardt iterative algorithm to maximize model likelihood and estimate parameters. We assumed a constant sex effect across

the latent classes and a linear time variable (age at pulmonary function measurement) that varied across classes. We specified a random intercept and slope for each participant in order to account for any correlation between repeated measures. We set the function to use a non-structured variance-covariance matrix for the random effects that is constant over the latent classes. In addition, we fit models with and without an autoregressive process to model the correlation between subject observations. Lastly, the function's default starting values were used for the estimation process. We also explored the sensitivity of final parameter estimates and the likelihood to different starting values. While some starting values lead to only local maxima, our reported results were consistently observed and were the observed maximum value of the likelihood. All models converged within 500 iterations.

To compute maximally attained FEV_1 and FVC, we used reference equations from Hankinson and colleagues to predict the maximum values for each participant.[3] The equations are maximized at age 18 for females and age 20 for males.[3] Other inputs include the participant's self-reported race and maximum measured height. These participant-specific values were subsequently used as denominators to compute percent maximum predicted (pmp) FEV_1 and FVC, respectively. Linear mixed models were used to model pmpFEV_1 and pmpFVC values as functions of class, spirometry measurement age, and their interaction (fixed effects), with a participant specific intercept (random effect) (Figures E1 and E2).

We utilized pre-bronchodilator spirometry measurements in order to strengthen the

generalizability of our results, as we wanted to include as many participants as possible in our analyses. There are significantly more missing data for post-bronchodilator spirometry compared to pre-bronchodilator measurements (e.g. 110 participants out of the 599 included in the primary analyses do not have post-bronchodilator measurements at age 32 years). In this general population cohort study of healthy individuals, not all participants and/or their parents who provided consent during the initial surveys were willing to undergo bronchodilator administration. Moreover, because of the duration of the study, many participants have moved away from the research center, prompting more home visits, and study nurses only perform post-bronchodilator spirometry in states where they are licensed to administer medications.

Notably, this approach focusing on pre-bronchodilator lung function is consistent with other published studies examining lung function trajectories. For example, pre-bronchodilator spirometry was used by Lange and colleagues in their landmark study that addressed the contributions of different lung function trajectories to the development of COPD.[4] Pre-bronchodilator spirometry was also used by McGeachie *et al* in their analysis of lung function data trajectories from the Childhood Asthma Management Program (CAMP) cohort.[5]

Results Supplement

The AIC criterion indicated that models with autoregressive error structure were preferred. Further, a model with two latent classes provided better fit than a model with

a single class (see Table 3), and a model with three classes was preferred to two classes. Model comparison using -2 log likelihood ratio also indicated that two latent classes were preferred to one (χ^2_3 = 17.1, p=0.0007), and that three classes were preferred to two (χ^2_3 = 13.9, p=0.003). Because only seven participants (~1%) were classified to the third group, we proceeded using the two-class model.

REFERENCES

1. Verbeke G, E L. A linear mixed-effects model with heterogeneity in the random-effects population. *J Am Stat Assoc* 1996; 91: 217-21.
2. lcmm: extended mixed models using latent classes and latent processes. R package. Accessed 2015, at http://CRAN.R-project.org/package=lcmm.
3. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179-187.
4. Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, Guerra S, Marott JL, Martinez FD, Martinez-Camblor P, Meek P, Owen CA, Petersen H, Pinto-Plata V, Schnohr P, Sood A, Soriano JB, Tesfaigzi Y, Vestbo J. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *New Engl J Med* 2015; 373(2): 111-22.
5. McGeachie MJ, Yates KP, Zhou X, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, Cho MH, Croteau-Chonka DC, Castaldi PJ, Jain G, Sanyal A, Zhan Y, Lajoie BR, Dekker J, Stamatoyannopoulos J, Covar RA, Zeiger RS, Adkinson NF, Williams PV, Kelly HW, Grasemann H, Vonk JM,

Koppelman GH, Postma DS, Raby BA, Houston I, Lu Q, Fuhlbrigge AL, Tantisira KG, Silverman EK, Tonascia J, Weiss ST, Strunk RC for the CAMP Research Group. Patterns of growth and decline in lung function in persistent childhood asthma. *N Engl J Med* 2016; 374: 1842-52.

Table E1. Enrollment characteristics of CRS participants with at least two lung function measurements included in analysis compared to those excluded from analysis

Enrollment characteristics	Included participants	Excluded participants	p
Total N	599	647	
Gender Male	293 (48.9%)	320 (49.5%)	0.85
Race/ethnicity Caucasian Mexican American African American	381 (66.4%) 175 (30.5%) 18 (3.1%)	440 (71.7%) 145 (23.6%) 29 (4.7%)	0.02
Birthweight (g)	3487 (470)	3424 (491)	0.02
Parental age (yr) Maternal Paternal	27.7 (4.5) 30.1 (5.9)	26.8 (4.8) 29.2 (5.6)	0.001 0.005
Parental asthma* Maternal Paternal	64 (10.9%) 66 (11.7%)	63 (11.1%) 66 (12.4%)	0.89 0.74
In-utero smoke exposure† Maternal Paternal	88 (14.7%) 165 (27.9%)	132 (20.5%) 220 (34.7%)	0.007 0.01
Parental education Maternal ≤12 yr Paternal ≤12 yr	159 (26.5%) 151 (25.7%)	235 (36.6%) 210 (33.4%)	<0.001 0.003

*Ever physician diagnosis of asthma. †Active parental smoking at birth. Categorical data are shown as n (%), and continuous data are show as mean (sd).

Table E2. Akaike information criteria (AIC) for models fit with various latent classes with or without autoregression (AR) including participants with at least two lung function measurements (n=599)

Latent Classes	AR	AIC
1	No	12825.34
1	Yes	12752.49
2	No	12831.34
2	Yes	12741.44
3	No	12793.76
3	Yes	12733.96

Table E3. Probability of lung function trajectory class membership based on two-class model including participants with at least two lung function measurements (n=599)

	Low Trajectory	Normal Trajectory
Probability ≥ 0.5	56 (100%)	543 (100%)
Probability ≥ 0.6	42 (75.0%)	519 (95.6%)
Probability ≥ 0.7	33 (58.9%)	495 (91.2%)
Probability ≥ 0.8	25 (44.6%)	461 (84.9%)
Probability ≥ 0.9	15 (26.8%)	408 (75.1%)

Data are shown as n (%).

Table E4. Regression coefficients for FEV_1/FVC ratio fixed effects model with two latent classes including participants with at least two lung function measurements (n=599)

	β	Std error	Wald statistic	p
Normal Trajectory				
Intercept (age 11)	85.59	0.702	124.88	<0.001
Slope	-0.19	0.018	-10.59	<0.001
Low Trajectory				
Intercept (age 11)	76.49	1.777	43.94	<0.001
Slope	-0.15	0.063	-2.30	0.022
Gender				
Male	-2.74	0.430	-6.37	<0.001

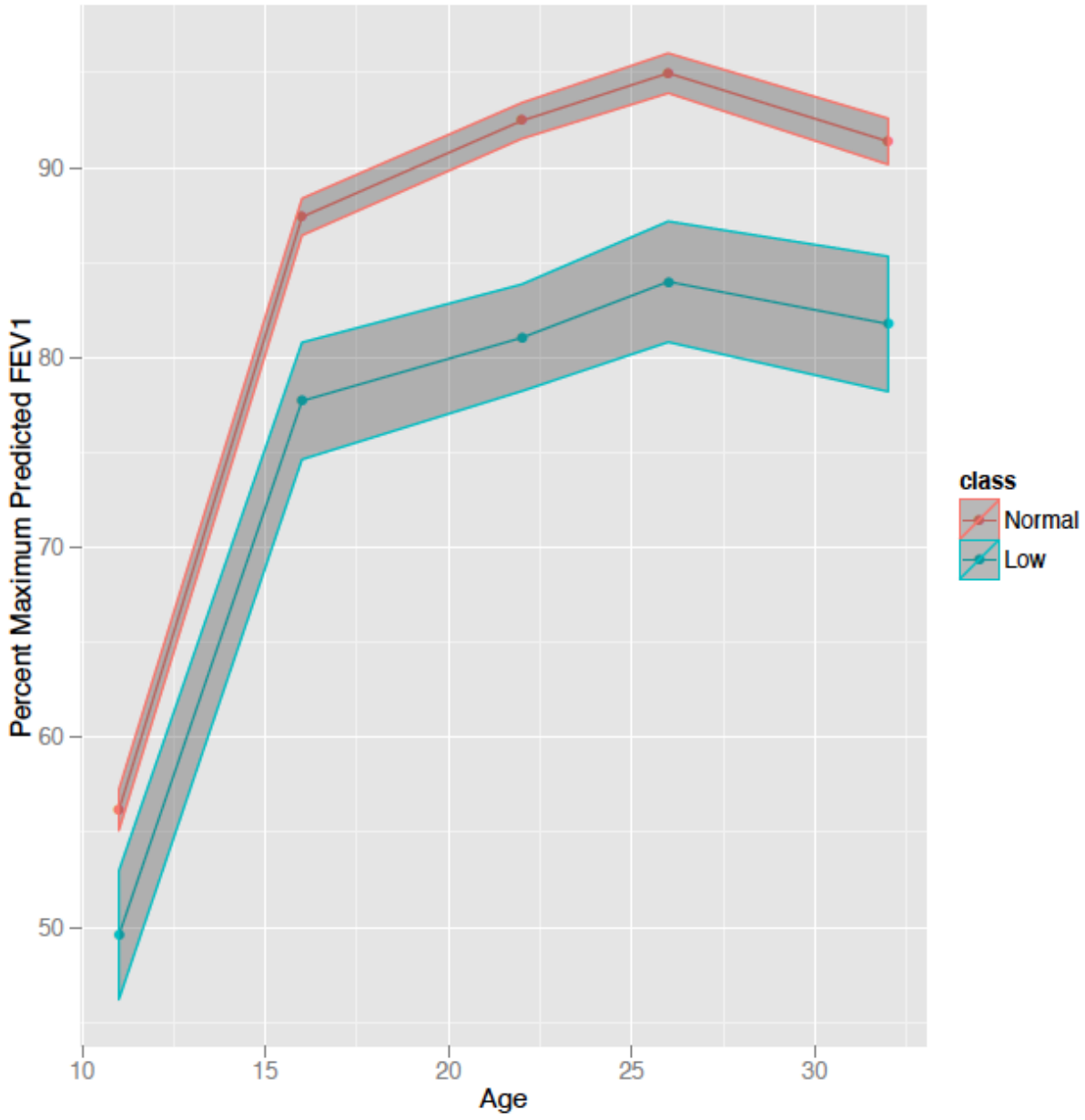
Table E5. Regression coefficients for FEV_1/FVC ratio fixed effects model with two latent classes including participants with at least three lung function measurements (n=480)

	β	Std error	Wald statistic	p
Normal Trajectory				
Intercept (age 11)	85.82	0.788	111.69	<0.001
Slope	-0.20	0.020	-9.63	<0.001
Low Trajectory				
Intercept (age 11)	76.73	1.623	48.12	<0.001
Slope	-0.12	0.055	-2.21	0.027
Gender				
Male	-2.64	0.480	-5.49	<0.001

Table E6. Regression coefficients for FEV_1/FVC ratio fixed effects model with two latent classes including participants with at least two lung function measurements adjusted for active physician-diagnosed (MD Dx) asthma (n=599)

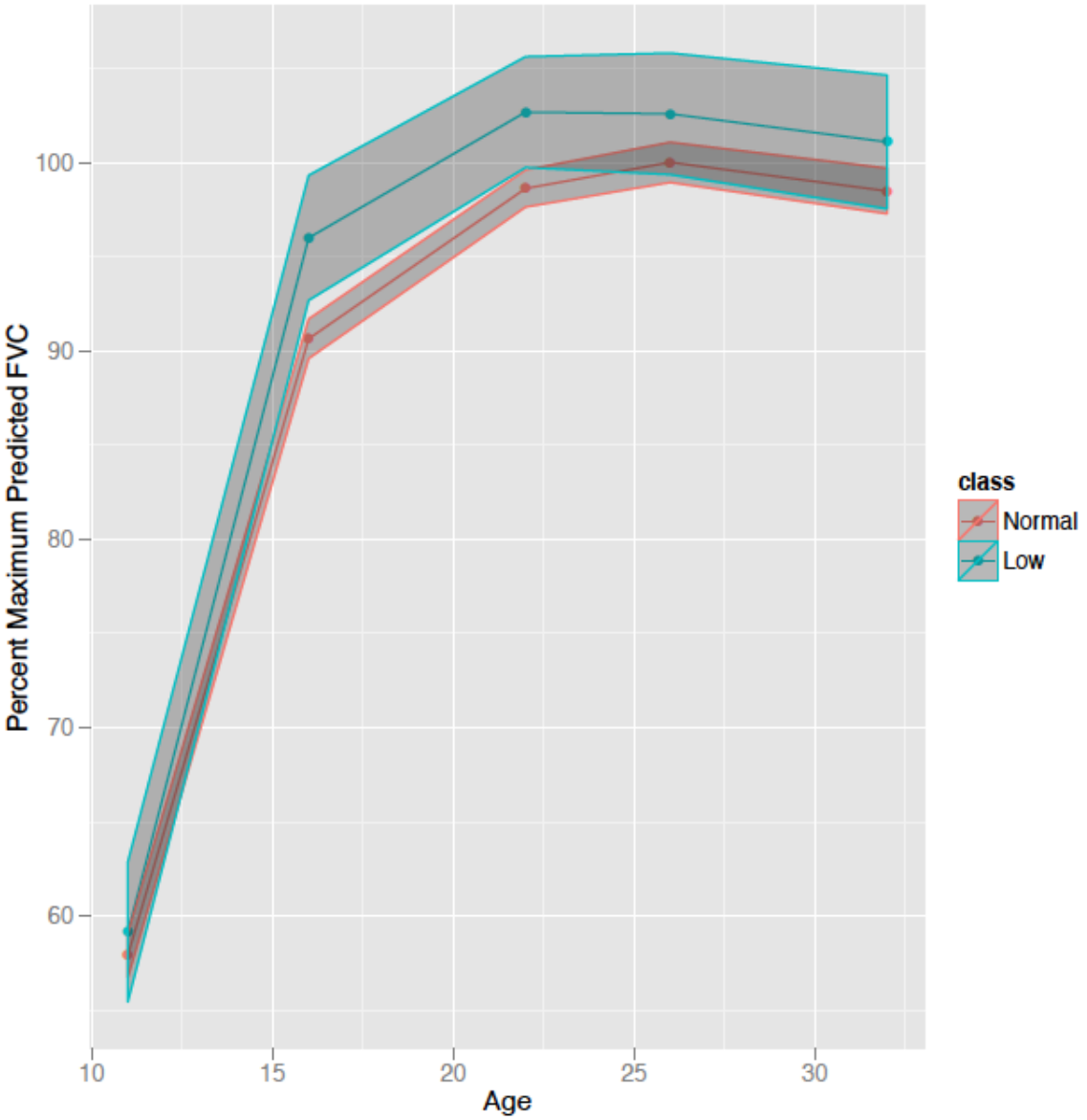
	β	Std error	Wald statistic	p
Normal Trajectory				
Intercept (age 11)	86.10	0.846	104.35	<0.001
Slope	-0.20	0.021	-9.27	<0.001
Low Trajectory				
Intercept (age 11)	77.65	1.580	50.05	<0.001
Slope	-0.13	0.054	-2.45	0.014
Gender				
Male	-2.61	0.482	-5.41	<0.001
Asthma				
Active symptoms, MD Dx	-1.16	0.335	-3.47	<0.001

Figure E1. Percentage of maximum predicted FEV_1 achieved for CRS participants assigned to normal and low lung function trajectory classes



Each point is the estimated mean and 95% confidence interval from the linear mixed model. Red indicates normal lung function class and cyan indicates low lung function class.

Figure E2. Percentage of maximum predicted FVC achieved for CRS participants assigned to normal and low lung function trajectory classes



Each point is the estimated mean and 95% confidence interval from the linear mixed model. Red indicates normal lung function class and cyan indicates low lung function class.